

Predicting Spread of Fake News on Social Media

Pedram Hosseini; Mona Diab; David A. Broniatowski
 {phosseini, mtidiab, broniatowski}@gwu.edu



THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

Introduction

%75 of Americans who recognized a *fake news* story from the US 2016 presidential election still viewed the story as accurate.



why?



Current methods to study and understand fake news phenomenon:

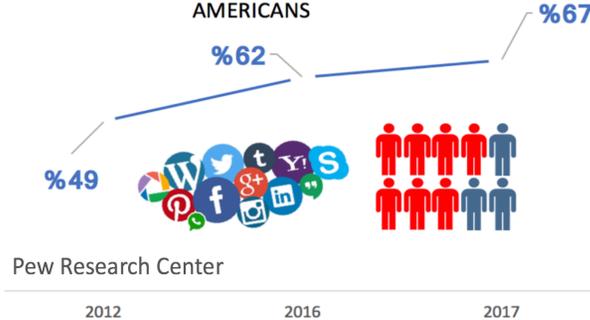
- Using Artificial Intelligence and machine learning models to detect fake news on social media and online platforms
- Statistical analysis of fake news and using network-based techniques to understand how misinformation spreads

But is it enough? **NO!**

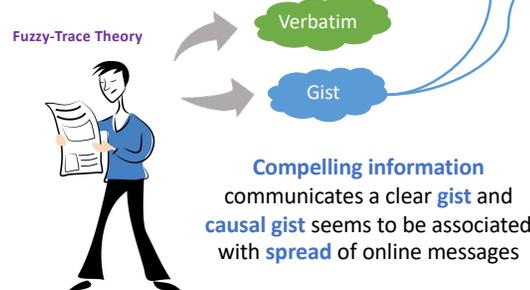
What we do



SOCIAL MEDIA AS MAIN SOURCE OF NEWS FOR AMERICANS



Why AI + Psychology



- **Gist**: subjective, but meaningful, interpretation
- **Verbatim**: objective, but decontextualized, facts

We have demonstrated that there is a correlation between **semantic/causal coherence** and **spread of fake/real news articles** on social media [1]

Experiments and results

Statistics of *FakeNewsNet* [2] dataset after pre-processing

Fact checker	Fake	Real	Total
PolitiFact	114	117	231
Buzzfeed	89	90	179
Total			410

Components	Description	Estimate	Std.error	t statistic
RC58	Number of syllables	0.11	0.03	3.56 ***
Truth labels	Fake/Real labels	0.32	0.09	3.54 ***
RC14	Preposition phrase density	-0.10	0.03	-3.27 **
RC16	Adverbial phrase density	-0.09	0.03	-3.02 **
RC2	Argument overlap, all sentences	0.09	0.03	2.86 **
RC24	First person plural pronoun incidence	0.09	0.03	2.80 **
RC72	Referential cohesion	-0.07	0.03	-2.42 *
RC7	Number of words	0.08	0.03	2.41 *
RC11	Number of letters	-0.07	0.03	-2.33 *
RC47	LSA overlap, adjacent paragraphs	-0.07	0.03	-2.22 *
RC31	Third person plural pronoun	-0.07	0.03	-2.18 *
RC66	Stem overlap, adjacent sentences	0.07	0.03	2.14 *
RC94	Hyponymy for nouns and verbs	-0.06	0.03	-2.05 *
RC30	Temporal connectives incidence	0.06	0.03	2.02 *
RC39	Word frequency for content words	-0.06	0.03	-2.00 *
RC37	Noun phrase density	-0.06	0.03	-1.98 *
(Intercept)	-	1.52	0.13	11.95

Table 1. Note.***=p < 0.001, ** = p < 0.01, * = p < 0.05. "RCs" are the PCA components. "Description" contains the description of the Coh-Metrix index with the highest absolute loading value in the corresponding PCA component.

Conclusion and Future work

- There are signals that show *more causally coherent stories, including mis-/disinformation, are more likely to be shared online.*
- We are using *Natural Language Processing (NLP)* to implement methods to better measure causal coherence for text.

References

- [1] Hosseini, P., Diab, M., and Broniatowski DA.: "Does Causal Coherence Predict Online Spread of Social Media?." SBP-BRiMS. Springer, (2019)
- [2] Shu, K., Wang, S., Liu, H.: "Exploiting tri-relationship for fake news detection." arXiv preprint arXiv:1712.07709 (2017)



Interested in #FakeNews related updates?
 Follow me @PedramHosseini on Twitter.